Bounds on Thresholds Related to Maximum Satisfiability of Regular Random Formulas

Vishwambhar Rathi*†, Erik Aurell*‡, Lars Rasmussen*†, Mikael Skoglund*†
{vish@, eaurell@, lars.rasmussen@ee, skoglund@ee}.kth.se
*School of Electrical Engineering
†KTH Linnaeus Centre ACCESS
KTH-Royal Institute of Technology, Stockholm, Sweden
‡Dept. Information and Computer Science, TKK-Helsinki University of Technology, Espoo, Finland

***Abstract*—We consider the regular balanced model of formula generation in conjunctive normal form (CNF) introduced by Boufkhad, Dubois, Interian, and Selman. We say that a formula is p-satisfying if there is a truth assignment satisfying $1-2^{-k}+p2^{-k}$ fraction of clauses. Using the first moment method we determine upper bound on the threshold clause density such that there are no p-satisfying assignments with high probability above this upper bound. There are two aspects in deriving the lower bound using the second moment method. The first aspect is, given any $p \in (0,1)$ and k, evaluate the lower bound on the threshold. This evaluation is numerical in nature. The second aspect is to derive the lower bound as a function of p for large enough k. We address the first aspect and evaluate the lower bound on the p-satisfying threshold using the second moment method. We observe that as k increases the lower bound seems to converge to the asymptotically derived lower bound for uniform model of formula generation by Achlioptas, Naor, and Peres.**

I. Regular Formulas and Motivation

A literal of a boolean variable is the variable itself or its negation. A clause is a disjunction (OR) of k literals. A formula is a conjunction (AND) of a finite set of clauses. A k-SAT formula is a formula where each clause is a disjunction of k literals. A *legal* clause is one in which there are no repeated or complementary literals. Using the terminology of [5], we say that a formula is *simple* if it consists of only legal clauses. A *configuration* formula is not necessarily legal. A satisfying (SAT) assignment of a formula is a truth assignment of variables for which the formula evaluates to true i.e. all the clauses evaluates to true. We denote the number of variables by n, the number of clauses by m, and the clause density i.e. the ratio of clauses to variables by $\alpha = \frac{m}{n}$. We denote the binary entropy function by $h(x) \triangleq -x\ln(x) - (1-x)\ln(1-x)$, where the logarithm is the natural logarithm.

The popular, uniform k-SAT model generates a formula by selecting uniformly and independently m-clauses from the set of all $2^k\binom{n}{k}$ k-clauses. In this model, the literal degree can vary. We are interested in the model where the literal degree is constant, which was introduced in [5]. Suppose each literal has degree r. Then $2nr = km$, which gives $\alpha = 2r/k$. Hence α can only take values from a discrete set of possible values. This problem can be circumvented by allowing each literal to take two possible values for a degree. However, in this paper we consider the case where all the literals have degree r due to space restriction. A formula is represented by a bipartite graph. The left vertices represent the literals and right vertices represent the clauses. A literal is connected to a clause if it appears in the clause. There are $k\alpha n$ edges coming out from all the literals and $k\alpha n$ edges coming out from the clauses. We assign the labels from the set $\mathcal{E} = \{1,\ldots,k\alpha n\}$ to edges on both sides of the bipartite graph. In order to generate a formula, we generate a random permutation Π on $\mathcal{E}$. Now we connect an edge i on the literal node side to an edge $\Pi(i)$ on the clause node side. This gives rise to a regular random k-SAT formula. Note that not all the formulas generated by this procedure are simple. However, it was shown in [5] that the threshold is same for this collection of formulas and the collection of simple formulas. Thus, we can work with the collection of configuration formulas generated by this procedure. Note that this procedure is similar to the procedure of generating regular LDPC codes [11].

The regular random k-SAT formulas are of interest because such instances are computationally harder than the uniform k-SAT instances. This was experimentally observed in [5], where the authors also derived upper and lower bounds on the satisfiability threshold for regular random 3-SAT. In [9], upper bound on the satisfiability threshold for any $k \geq 3$ was derived using the first moment method. It was shown that as k increases, the upper bound converge to the corresponding bounds on the threshold of the uniform model [1], [3].

For uniform model, in a series of breakthrough papers, Achlioptas and Peres in [3] and Achlioptas and Moore in [1], derived almost matching lower bounds with the upper bound by carefully applying the second moment method to balanced satisfying assignments. In [1], based on their belief that the simple application of second moment method should work for symmetric problems, Achlioptas and Moore posed the question of its success for regular random k-SAT. In an attempt to answer this question, the lower bound using the second moment method was evaluated in [9]. As the evaluation of the lower bound was numerical in nature (though exact), it was observed that the lower bound also converges to the corresponding lower bound for the uniform model as k increases.

In this work we are interested in the maximum satisfiability problem over regular random formulas. We say that a formula

is p-satisfying if there is a truth assignment satisfying $c(p) \triangleq 1-2^{-k}+p2^{-k}$ fraction of its clauses, $p \in (0,1)$. The number of p-satisfying truth assignments is denoted by $N(n,\alpha,p)$. We define the following quantities related to p-satisfiability:

$$\begin{aligned} \alpha(p) &\triangleq \sup\{\alpha : \text{A regular random formula, is } p\text{-satisfiable w.h.p.}\}, \quad (1) \\ \alpha^*(p) &\triangleq \inf\{\alpha : \text{A regular random formula, is not } p\text{-satisfiable w.h.p.}\}. \quad (2) \end{aligned}$$

Note that $\alpha(p) \leq \alpha^*(p)$. In [2], for the uniform model Achlioptas, Naor, and Peres derived lower bound on $\alpha(p)$ which almost matches with the upper bound on $\alpha^*(p)$ derived via the first moment method. The lower bound was obtained by a careful application of the second moment method.

We derive upper bound on $\alpha^*(p)$ by applying the first moment method to $N(n,\alpha,p)$. The obtained upper bound matches with the corresponding bound for the uniform model. We evaluate a lower bound on $\alpha(p)$ by applying the second moment method to $N(n,\alpha,p)$. We observe that for increasing k the lower bound seems to converge to the corresponding bound for the uniform model. In the next section, we obtain upper bound on $\alpha^*(p)$ by the first moment method.

Due to space limitations, some of the arguments are accompanied by short explanations. Further details can be found in the forthcoming journal submission [10].

II. Upper Bound on Threshold via First Moment

Let X be a non-negative integer-valued random variable and $E(X)$ be its expectation. Then the first moment method gives: $\mathrm{P}(X>0) \leq E(X)$. Note that by choosing X to be the number of solutions of a random formula, we can obtain an upper bound on the threshold $\alpha^*(p)$ beyond which no p-satisfying solution exists with probability one. This upper bound corresponds to the largest value of α at which the average number of p-satisfying solutions goes to zero as n tends to infinity. In the following lemma, we derive the first moment of $N(n,\alpha,p)$ for regular random k-SAT for $k \geq 2$.

Lemma 2.1: Let $N(n,\alpha,p)$ be the number of p-satisfying assignments for a random regular k-SAT formula. Then[1],

$$\begin{aligned} E(N(n,\alpha,p)) &= 2^n \binom{\alpha n}{c(p)\alpha n} \frac{\left(\left(\frac{k\alpha n}{2}\right)!\right)^2}{(k\alpha n)!} \times \\ &\quad \mathrm{coef}\left(\left(\frac{s(x)}{x}\right)^{c(p)\alpha n}, x^{\left(\frac{k}{2}-c(p)\right)\alpha n}\right), \quad (3) \end{aligned}$$

where $s(x) = (1+x)^k - 1$ and $\mathrm{coef}\left(\left(\frac{s(x)}{x}\right)^{c(p)\alpha n}, x^{\left(\frac{k}{2}-c(p)\right)\alpha n}\right)$ denotes the coefficient of $x^{\left(\frac{k}{2}-c(p)\right)\alpha n}$ in the expansion of $\left(\frac{s(x)}{x}\right)^{c(p)\alpha n}$.

Proof: Due to symmetry of the formula generation, any assignment of variables has the same probability of being p-satisfying. This implies

$$E(N(n,\alpha,p)) = 2^n \mathrm{P}(X = \{0,\ldots,0\} \text{ is } p\text{-satisfying.}).$$

[1]We assume that $k\alpha n$ is an even integer.

The probability of the all-zero vector being p-satisfying is given by

$$\mathrm{P}(X = \{0,\ldots,0\} \text{ is } p\text{-satisfying}) = \frac{\text{Number of formulas for which } X = \{0,\ldots,0\} \text{ is } p\text{-satisfying}}{\text{Total number of formulas}}.$$

The total number of formulas is given by $(k\alpha n)!$. The total number of formulas for which the all-zero assignment is a p-satisfying is given by

$$\binom{\alpha n}{c(p)\alpha n}\left(\left(\frac{k\alpha n}{2}\right)!\right)^2 \mathrm{coef}\left(s(x)^{c(p)\alpha n}, x^{\frac{k\alpha n}{2}}\right).$$

The binomial term corresponds to choosing $c(p)$ fraction of clauses being satisfied by the all-zero assignment. The factorial terms correspond to permuting the edges among true and false literals. Note that there are equal numbers of true and false literals. The generating function $s(x)$ corresponds to placing at least one positive literal in a clause. With these results and observing that $\mathrm{coef}\left(s(x)^{c(p)\alpha n}, x^{\frac{k\alpha n}{2}}\right) = \mathrm{coef}\left((s(x)/x)^{c(p)\alpha n}, x^{\left(\frac{k}{2}-c(p)\right)\alpha n}\right)$, we obtain (3). ∎

Remark: The computation of the first moment of $N(n,\alpha,p)$ is similar to the computation of the first moment for weight distribution of regular LDPC ensembles. There is large body of work dealing with first moment of weight distribution. So, we refer to [11] and the references there in.

We now state the Hayman method to approximate the coef-term which is asymptotically correct [11].

Lemma 2.2 (Hayman Method): Let $q(y) = \sum_i q_i y^i$ be a polynomial with non-negative coefficients such that $q_0 \neq 0$ and $q_1 \neq 0$. Define

$$a_q(y) = y\frac{dq(y)}{dy}\frac{1}{y}, \quad b_q(y) = y\frac{da_q(y)}{dy}. \quad (4)$$

Then,

$$\mathrm{coef}(q(y)^n, y^{\omega n}) = \frac{q(y_\omega)^n}{(y_\omega)^{\omega n}\sqrt{2\pi n b_q(y_\omega)}}(1+o(1)), \quad (5)$$

where y_ω is the unique positive solution of the saddle point equation $a_q(y) = \omega$. The solution y_ω also satisfies

$$y_\omega = \inf_{y>0} \frac{q(y)^n}{y^{\omega n}}. \quad (6)$$

We now use Lemma 2.2 to compute the expectation of the total number of p-satisfying assignments.

Lemma 2.3: Let $N(n,\alpha,p)$ denote the total number of p-satisfying assignments of a regular random k-SAT formula. Let $t(x) = \frac{s(x)}{x}$, where $s(x)$ is defined in Lemma 2.1. Then,

$$\begin{aligned} E(N(n,\alpha,p)) &= \sqrt{\frac{k}{8\pi c(p)^2(1-c(p))b_t(x_k)\alpha n}}\, e^{n((1-k\alpha)\ln(2))} \\ &\quad e^{n\left(\alpha h(c(p))+c(p)\alpha\ln(t(x_k))-\left(\frac{k\alpha}{2}-\alpha c(p)\right)\ln(x_k)\right)}(1+o(1)), \quad (7) \end{aligned}$$

where x_k is the solution of $a_t(x) = \frac{k}{2c(p)} - 1$. The quantity $a_t(x)$ and $b_t(x)$ are defined according to (4).

Proof: Using Stirling's approximation for the binomial terms (see [11, p. 513]) and Hayman approximation for the coef term from Lemma 5 gives the desired result. ∎

In the following lemma we derive explicit upper bounds on the clause density for the existence of p-satisfying assignments.

Lemma 2.4 (Upper bound): Let $\alpha^*(p)$ be as defined in (2). Define $\alpha_u^*(p)$ to be the upper bound on $\alpha^*(p)$ obtained by the first moment method. Then,

$$\alpha^*(p) \leq \alpha_u^*(p) \triangleq \frac{2^k \ln(2)}{p+(1-p)\ln(1-p)}. \tag{8}$$

Proof: Using (6), we obtain the following upper bound on the exponent of $E(N(n,\alpha,p)))$ for any $x>0$,

$$\lim_{n\to\infty} \frac{\ln\left(E(N(n,\alpha))\right)}{n} \leq (1-k\alpha)\ln(2)+\alpha h(c(p))+ c(p)\alpha\ln(t(x))-\alpha\left(\frac{k}{2}-c(p)\right)\ln(x). \tag{9}$$

We substitute $x=1$ in (9) and obtain the following upper bound on the p-satisfiability threshold,

$$\alpha^*(p) \leq \frac{\ln(2)}{k\ln(2)-h(c(p))-c(p)\ln(2^k-1)}. \tag{10}$$

We complete the proof by showing that the denominator in (10) is lower bounded by $2^{-k}(p+(1-p)\ln(1-p))$. This can be easily shown by considering their difference and then showing its positivity. ∎

Note that the upper bound coincides with the upper bound derived for uniform formulas in [2]. In the next section we use the second moment method to obtain lower bound on $\alpha(p)$ defined in (1).

III. Second Moment

In [2], almost matching lower bounds on the p-satisfiability threshold of uniform formulas were derived using the second moment method. The second moment method is governed by the following equation

$$\mathrm{P}(X>0) \geq \frac{E(X)^2}{E(X^2)}, \tag{11}$$

where X is a non-negative random variable. Before we use the second moment method, we present the following theorem from [6] and its corollary given in [2]. Proof of both theorem and the corollary are identical for the uniform model and the regular model.

Theorem 3.1 ([6]): Let $U_k(n,\alpha)$ be the maximum number of clauses satisfied (over all the truth assignments) of a regular random formula with n variables and m clauses. Then

$$\mathrm{P}\left(|U_k(n,\alpha)-E\left(U_k(n,\alpha)\right)|>t\right) < 2\exp\left(-\frac{2t^2}{\alpha n}\right).$$

Corollary 3.1 ([2]): Assume that there exists $c=c(k,p,r)$ such that for n large enough, a regular random formula is p-satisfiable with probability greater than n^{-c}. Then a regular random formula is p'-satisfiable with high probability for every constant $p'<p$.

We now apply the second moment method to $N(n,\alpha,p)$. The calculation for $p=1$ i.e. for number of satisfying assignments was done in [9]. Our computation of the second moment is inspired by the computation of the second moment for the weight and stopping set distributions of regular LDPC codes in [7], [8] (see also [4]). We compute the second moment in the next lemma.

Lemma 3.1: Consider regular random satisfiability formulas with literal degree r. Then the second moment of $N(n,\alpha,p)$ is given by

$$E\left(N(n,\alpha,p)^2\right) = \sum_{i=0}^{n} \sum_{j=\alpha n\left(c(p)-(1-p)2^{-k}\right)}^{c(p)\alpha n} 2^n \binom{n}{i}\binom{\alpha n}{c(p)\alpha n} \binom{c(p)\alpha n}{j}\binom{\alpha n(1-c(p))}{c(p)\alpha n-j} \frac{\left((r(n-i))!\right)^2\left((ri)!\right)^2}{(k\alpha n)!} \mathrm{coef}\left(f(x_1,x_2,x_3)^j (s(x_1)s(x_3))^{c(p)\alpha n-j}, (x_1x_3)^{r(n-i)}x_2^{ri}\right), \tag{12}$$

where the generating function $f(x_1,x_2,x_3)$ is given by

$$f(x_1,x_2,x_3) = (1+x_1+x_2+x_3)^k-(1+x_1)^k-(1+x_3)^k+1. \tag{13}$$

The generating function $s(x)$ is defined as $s(x) \triangleq (1+x)^k-1$, which is same as defined in Lemma 2.1.

Proof: For truth assignments X and Y, define the indicator variable $\mathbb{1}_{XY}$ which evaluates to 1 if the truth assignments X and Y are p-satisfying. Then,

$$\begin{aligned} E(N(n,\alpha,p)^2) &= \sum_{X,Y\in\{0,1\}^n} E\left(\mathbb{1}_{\mathbf{XY}}\right), \\ &= 2^n \sum_{Y\in\{0,1\}^n} \mathrm{P}\left(\mathbf{0} \text{ and } Y \text{ are } p\text{-satisfying}\right). \end{aligned}$$

Due to the symmetry in regular formula generation, the number of formulas for which both X and Y are p-satisfying depends only on the number of variables on which X and Y agree. This explains the last simplification where we fix X to be the all-zero vector.

We want to evaluate the probability of the event that the truth assignments $\mathbf{0}$ and Y p-satisfy a randomly chosen regular formula. This probability depends only on the *overlap*, i.e., the number of variables where the two truth assignments agree. Thus for a given overlap i, we can fix Y to be equal to zero in the first i variables and equal to 1 in the remaining variables i.e. $Y=\{\underbrace{0,\cdots,0}_{i \text{ times}},\underbrace{1,\cdots,1}_{n-i \text{ times}}\}$. This gives,

$$E(N(n,\alpha,p)^2) = \sum_{i=0}^{n} 2^n \binom{n}{i} \mathrm{P}\left(\mathbf{0} \text{ and } Y \text{ are } p\text{-satisfying}\right). \tag{14}$$

In order to evaluate the probability that both $\mathbf{0}$ and Y are p-satisfying, define $\mathcal{C}=\{1,\cdots,\alpha n\}$ to be the set of clauses and $C_{\mathbf{0}}$ and C_Y to be the set of clauses satisfied by $\mathbf{0}$ and Y respectively. Clearly, $|C_{\mathbf{0}}|=|C_Y|=c(p)\alpha n$. Then,

$$\mathrm{P}\left(\mathbf{0} \text{ and } Y \text{ are } p\text{-satisfying}\right) = \sum_{C_{\mathbf{0}},C_Y\subset\mathcal{C}} \mathrm{P}\left(\mathbf{0} \text{ only satisfies } C_{\mathbf{0}} \text{ and } Y \text{ only satisfies } C_Y\right). \tag{15}$$

Again from the symmetry of the regular formula generation, we fix $C_{\mathbf{0}} = \{1,\cdots,c(p)\alpha n\}$. For $|C_{\mathbf{0}} \cap C_Y| = j$, we fix $C_Y = \{1,\cdots,j,c(p)\alpha n+1,\ldots,2c(p)\alpha n - j\}$. This gives,

$$\mathrm{P}(\mathbf{0} \text{ and } Y \text{ are } p\text{-satisfying}) = \binom{\alpha n}{c(p)\alpha n} \sum_{j=\alpha n\left(c(p)-(1-p)2^{-k}\right)}^{c(p)\alpha n} \binom{c(p)\alpha n}{j}\binom{\alpha n(1-c(p))}{c(p)\alpha n - j} \times \mathrm{P}(\mathbf{0} \text{ only satisfies } C_{\mathbf{0}} \text{ and } Y \text{ only satisfies } C_Y). \quad (16)$$

Note that $j \geq \alpha n(c(p)-(1-p)2^{-k})$ as $|C_{\mathbf{0}}| + |C_Y| - |C_{\mathbf{0}} \cap C_Y| \leq \alpha n$. For a given overlap i between $\mathbf{0}$ and Y, we observe that there are four different types of edges connecting the literals and the clauses. There are $r(n-i)$ **type 1** edges which are connected to true literals w.r.t. the $\mathbf{0}$ truth assignment and false w.r.t. to the Y truth assignment. The ri **type 2** edges are connected to true literals w.r.t. both the truth assignments. There are $r(n-i)$ **type 3** edges which are connected to false literals w.r.t. the $\mathbf{0}$ truth assignment and true literals w.r.t. to the Y truth assignment. The ri **type 4** edges are connected to false literals w.r.t. both the truth assignments. Let $f(x_1,x_2,x_3)$ be the generating function counting the number of possible edge connections to a clause such that the clause is satisfied by both $\mathbf{0}$ and Y. In $f(x_1,x_2,x_3)$, the power of x_i gives the number of edges of type i, $i \in \{1,2,3\}$. A clause is satisfied by both $\mathbf{0}$ and Y if it is connected to at least one type 2 edge, else it is connected to at least one type 1 and at least one type 3 edge. Then the generating function $f(x_1,x_2,x_3)$ is given as in (13). Using this, we obtain

$$\mathrm{P}(\mathbf{0} \text{ only satisfies } C_{\mathbf{0}} \text{ and } Y \text{ only satisfies } C_Y) = \frac{((r(n-i))!)^2\,((ri)!)^2}{(k\alpha n)!} \times \mathrm{coef}\left(f(x_1,x_2,x_3)^j (s(x_1)s(x_3))^{c(p)\alpha n-j}, (x_1x_3)^{r(n-i)}x_2^{ri}\right), \quad (17)$$

where $s(x_1)$ is the generating function for clauses satisfied by $\mathbf{0}$ and not satisfied by Y (similarly we define $s(x_3)$). The term $(k\alpha n)!$ is the total number of formulas. Consider a given formula which is satisfied by both truth assignments $\mathbf{0}$ and Y. If we permute the positions of type 1 edges on the clause side, we obtain another formula having $\mathbf{0}$ and Y as solutions. The argument holds true for the type i edges, $i \in \{2,3,4\}$. This explains the term $(r(n-i))!$ in (17) which corresponds to permuting the type 1 edges (it is squared because of the same contribution from type 3 edges). Similarly, $(ri)!^2$ corresponds to permuting type 2 and type 4 edges. As $|C_{\mathbf{0}} \cap C_Y| = j$, there are j clauses which satisfied by both $\mathbf{0}$ and Y. This explains the factor $f(x_1,x_2,x_3)^j$ in the coef term. There are $|C_{\mathbf{0}} \backslash (C_{\mathbf{0}} \cup C_Y)|) = |C_Y \backslash (C_{\mathbf{0}} \cup C_Y)| = c(p)\alpha n - j$ clauses which are satisfied by $\mathbf{0}$ (resp. Y) and not satisfied by Y (resp. $\mathbf{0}$). This explains the factor $(s(x_1)s(x_3))^{c(p)\alpha n-j}$ in the coef term. We complete the proof by substituting (17) in (16), then (16) in (14). ∎

In order to evaluate the second moment, we now present the multidimensional saddle point method in the next lemma. A detailed technical exposition of the multidimensional saddle point method can be found in Appendix D of [11].

Theorem 3.2: Let $\underline{i} := (r(n-i), ri, r(n-i))$, $\underline{x} = (x_1,x_2,x_3)$, and $0 < \lim_{n\to\infty} i/n < 1$. We define

$$g_{n,j}(\underline{x}) = f(\underline{x})^j (s(x_1)s(x_3))^{c(p)\alpha n-j},$$

where $f(\underline{x})$ and $s(x)$ are defined in Lemma 3.1. We define the normalizations $\eta \triangleq i/n$ and $\gamma \triangleq j/(\alpha n)$. Let $\underline{t} = (t_1,t_2,t_3)$ be a positive solution of the saddle point equations

$$a_g(\underline{x}) \triangleq \left\{\frac{x_i}{n}\frac{\partial \ln(g_{n,j}(\underline{x}))}{\partial x_i}\right\}_{i=1}^{3} = \{r(1-\eta), r\eta, r(1-\eta)\}. \quad (18)$$

Then $\mathrm{coef}\left(g_{n,j}(\underline{x}), \underline{x}^{\underline{i}}\right)$ can be approximated as ,

$$\mathrm{coef}\left(g_{n,j}(\underline{x}), \underline{x}^{\underline{i}}\right) = \frac{g_{n,j}(\underline{t})}{(\underline{t})^{\underline{i}}\sqrt{(2\pi n)^3\,|B_g(\underline{t})|}}(1+o(1)),$$

using the saddle point method for multivariate polynomials, where $B_g(\underline{x})$ is a 3×3 matrix whose elements are given by $B_{i,j} = x_j \frac{\partial a_{gi}(\underline{x})}{\partial x_j} = B_{j,i}$ and $a_{gi}(\underline{x})$ is the i^{th} coordinate of $a_g(\underline{x})$.

In the following theorem we derive lower bound on the p-satisfiability threshold by evaluating the second moment of $N(n,\alpha,p)$ with the help of Theorem 3.2.

Theorem 3.3: Consider regular random k-SAT formulas with literal degree r. Let $S(i,j)$ denote the $(i,j)^{\text{th}}$ summation term in (12). Define the normalization $\eta = i/n$ and $\gamma = j/(\alpha n)$. If $S(n/2, n\alpha c(p)^2)$ is the dominant term i.e. for $\eta \in [0,1], \gamma \in [(c(p)-2^{-k}(1-p)), c(p)], \eta \neq \frac{1}{2}$, and $\gamma \neq c(p)^2$

$$\lim_{n\to\infty}\frac{\ln\left(S\left(\frac{n}{2}, n\alpha c(p)^2\right)\right)}{n} > \lim_{n\to\infty}\frac{\ln(S(\eta n, \gamma\alpha n))}{n}, \quad (19)$$

then for some positive constants c, c'

$$\mathrm{P}(N(n,\alpha,p) > 0) \geq c' n^{-c}. \quad (20)$$

Let $r^*(p)$ be the largest literal degree for which $S(n/2, n\alpha c(p)^2)$ is the dominant term, i.e. (19) holds. Then the threshold $\alpha(p)$ defined in (1) is lower bounded by $\alpha(p) \geq \alpha_l^*(p) \triangleq \frac{2r^*(p)}{k}$.

Proof: Assuming (19) holds, then for n large enough

$$E\left(N(n,\alpha,p)^2\right) \leq (n+1)(\alpha(1-p)2^{-k}n+1)S\left(\frac{n}{2}, n\alpha c(p)^2\right). \quad (21)$$

From (12) and Theorem 3.2, the growth rate of $S(\eta n, \gamma\alpha n)$ is given by,

$$\begin{aligned} s(\eta,\gamma) \triangleq \lim_{n\to\infty}\frac{\ln(S(\eta n,\gamma\alpha n))}{n} &= (1-k\alpha)(\ln(2)+h(\eta)) \\ &+ \alpha h(c(p)) + \alpha c(p) h\left(\frac{\gamma}{c(p)}\right) + \alpha(1-c(p))h\left(\frac{c(p)-\gamma}{(1-c(p))}\right) \\ &+ \gamma\alpha\ln(f(t_1,t_2,t_3)) + \alpha(c(p)-\gamma)(\ln(s(t_1))+\ln(s(t_3))) \\ &- r(1-\eta)(\ln(t_1)+\ln(t_3)) - r\eta\ln(t_2), \quad (22) \end{aligned}$$

where t_1,t_2,t_3 is a positive solution of the saddle point equations as defined in Theorem 3.2,

$$a_g(\underline{t}) = \{r(1-\eta), r\eta, r(1-\eta)\}. \quad (23)$$

In order to compute the maximum exponent of the summation terms, we compute its partial derivatives with respect to η and γ and equate them to zero, which result in the following respective equations.

$$(1-k\alpha)\ln\frac{1-\eta}{\eta}+r\ln\left(\frac{t_1t_3}{t_2}\right) = 0, \quad (24)$$

$$\ln\left(\frac{(c(p)-\gamma)^2}{\gamma(1-2c(p)+\gamma)}\right)+\ln\left(\frac{f(t_1,t_2,t_3)}{s(t_1)s(t_3)}\right) = 0. \quad (25)$$

Note that the partial derivatives of t_1,t_2 and t_3 w.r.t. η and γ vanish because of the saddle point equations given in (23). Every positive solution (t_1,t_2,t_3) of (23) satisfies $t_1=t_3$ as (23) and $f(t_1,t_2,t_3)$ are symmetric in t_1 and t_3. If $\eta=1/2,\gamma=c(p)^2$ is a maximum, then the vanishing derivative in (24) and equality of t_1 and t_3 imply $t_2=t_1^2$. We substitute $\eta=1/2$, $t_1=t_3$, and $t_2=t_1^2$ in (23). This reduces (23) to the saddle point equation corresponding to the polynomial $t(x)$ defined in Lemma 2.3 whose solution is denote by x_k. By observing $f(x_k,x_k^2,x_k)=s(x_k)^2$, we have

$$S\left(\frac{n}{2},\alpha c(p)^2 n\right)=\frac{k^{3/2}}{32\pi^2 c(p)^2(1-c(p))^2\sqrt{|B_g(x_k,x_k^2,x_k)|}n^2}\times$$
$$e^{2n\left((1-k\alpha)\ln(2)+\alpha h(c(p))+\alpha c(p)\ln(s(x_k))-\frac{k\alpha}{2}\ln(x_k)\right)}(1+o(1)). \quad (26)$$

Using the relation that $t(x)=\frac{s(x)}{x}$, we note that the exponent of $S(n/2,\alpha c(p)^2 n)$ is twice the exponent of the first moment of the total number of solutions as given in (7). We substitute (26) in to (21). Then we use Lemma 2.3 and (21) in the second moment method:

$$\begin{aligned}\mathrm{P}(N(n,\alpha,p)>0) &\geq \frac{E(N(n,\alpha,p)^2)}{E(N(n,\alpha,p)^2)},\\ &\geq \frac{4\pi\sqrt{|B_g(x_k,x_k^2,x_k)|}}{b_t(x_k)\alpha^2\sqrt{k}n}(1+o(1)).\end{aligned}$$

Clearly, if the maximum of the growth rate of $S(\eta n,\gamma\alpha n)$ is not achieved at $\eta=1/2$ and $\gamma=c(p)^2$, then the lower bound given by the second moment method converges to zero exponentially fast. By using Corollary 3.1, we obtain the desired lower bound on $\alpha(p)$. This proves the theorem. ∎

In the next section we discuss the obtained lower and upper bounds on the p-satisfiability threshold.

IV. Bounds on Threshold and Discussion

In Table I, we compute the ratio of $\alpha_l^\star(p)$ and $\alpha_u^\star(p)$ for $p=0.1,\cdots,0.9$ and $k=3,6,12$, where $\alpha_l^\star(p)$ is the lower bound on $\alpha(p)$ obtained from Theorem 3.3 and $\alpha_u^*(p)$ is the upper bound on $\alpha^*(p)$ defined in Lemma 2.4. Note that the case $p=1$ was already solved in [9]. In order to apply the second moment method, we have to verify that $s(\eta,\gamma)$, defined in (22), attains its maximum at $\eta=\frac{1}{2},\gamma=c(p)^2$ over $[0,1]\times[c(p)-(1-p)2^{-k},c(p)]$. This requires that $\eta=\frac{1}{2},\gamma=c(p)^2$ is the only positive solution of the system of equations consisting of (23), (24) and (25) which corresponds to a maximum. The system of equations (23), (24), and (25) is equivalent to a system of polynomial equations. For small values of k and p close to one, we can solve this system of polynomial equations and verify the desired conditions. For larger values of k, the degree of monomials in (24) grows exponentially in k. Thus, solving (23), (24), and (25) becomes computationally difficult. However, $s(\eta,\gamma)$ can be easily computed as its computation requires solving only (23), where the maximum monomial degree is linear in k. Thus, the desired condition for maximum of $s(\eta,\gamma)$ at $\eta=\frac{1}{2},\gamma=c(p)^2$ can be verified numerically in an efficient manner.

p	$k=3$	$k=6$	$k=12$
0.1	0.252	0.717	0.977
0.2	0.258	0.720	0.979
0.3	0.272	0.738	0.980
0.4	0.281	0.755	0.981
0.5	0.295	0.765	0.983
0.6	0.308	0.782	0.986
0.7	0.325	0.801	0.988
0.8	0.344	0.822	0.990
0.9	0.402	0.855	0.993

TABLE I: *Value of the ratio* $\alpha_l^*(p)/\alpha_u^*(p)$.

From the Table I, we see that as k becomes larger the ratio $\alpha_l^*(p)/\alpha_u^*(p)$ gets closer to one. Our belief is that indeed as k becomes larger and larger, the ratio $\alpha_l^*(p)/\alpha_u^*(p)$ converges to one. Our main future goal is to derive explicit expression for $\alpha_l^*(p)$ as k becomes larger.

Acknowledgement

This work has been supported by Swedish research council (VR) through KTH Linnaeus center ACCESS.

References

[1] D. Achlioptas and C. Moore, *Random k-SAT: Two moments suffice to cross a sharp threshold*, SIAM J. COMPUT., 36 (2006), pp. 740–762.

[2] D. Achlioptas, A. Naor, and Y. Peres, *On the maximum satisfiability of random formulas*, Journal of the Association of Computing Machinary (JACM), 54 (2007).

[3] D. Achlioptas and Y. Peres, *The threshold for random k-SAT is* $2^k\ln(2)-O(k)$, Journal of the American Mathematical Society, 17 (2004), pp. 947–973.

[4] O. Barak and D. Burshtein, *Lower bounds on the spectrum and error rate of LDPC code ensembles*, in International Symposium on Information Theory, Adelaide, Australia, 2005.

[5] Y. Boufkhad, O. Dubois, Y. Interian, and B. Selman, *Regular random k-SAT: Properties of balanced formulas*, Journal of Automated Reasoning, (2005).

[6] A. Z. Broder, A. M. Frieze, and E. Upfal, *On the satisfiability and maximum satisfiability of random 3-CNF formulas*, Proc. 4th Annual Symposium on Discrete Algorithms, (1993), pp. 322–330.

[7] V. Rathi, *On the asymptotic weight and stopping set distributions of regular LDPC ensembles*, IEEE Trans. Inform. Theory, 52 (2006), pp. 4212–4218.

[8] ———, *Non-binary LDPC codes and EXIT like functions*, PhD thesis, Swiss Federal Institute of Technology (EPFL), Lausanne, 2008.

[9] V. Rathi, E. Aurell, L. Rasmussen, and M. Skoglund, *Bounds on thresholds of regular random k-SAT*. Accepted to International Conference on Theory and Applications of Satisfiability Testing (SAT) 2010.

[10] ———, *Satisfiability and maximum satisfiability of regular random k-sat: Bounds on thresholds*. in preparation for submission to IEEE Transactions on Information Theory.

[11] T. Richardson and R. Urbanke, *Modern Coding Theory*, Cambridge University Press, 2008.